

The Entropy Region for Three Gaussian Random Variables

Babak Hassibi and Sormeh Shadbakht

Department of Electrical Engineering

California Institute of Technology

Pasadena, CA 91125

hassibi,sormeh@caltech.edu

Abstract—Given n (discrete or continuous) random variables X_i , the $(2^n - 1)$ -dimensional vector obtained by evaluating the joint entropy of all non-empty subsets of $\{X_1, \dots, X_n\}$ is called an entropic vector. Determining the region of entropic vectors is an important open problem in information theory. Recently, Chan has shown that the entropy regions for discrete and continuous random variables, though different, can be determined from one another. An important class of continuous random variables are those that are vector-valued and jointly Gaussian. It is known that Gaussian random variables violate the Ingleton bound, which many random variables such as those obtained from linear codes over finite fields do satisfy, and they also achieve certain non-Shannon inequalities. In this paper we give a full characterization of the entropy region for three jointly-Gaussian vector-valued random variables and, rather surprisingly, show that the region is strictly smaller than the entropy region for three arbitrary random variables. However, we also show the following result. For any given entropic vector $h \in \mathcal{R}^7$, there exists a $\theta^* > 0$, such that for all $\theta \geq \theta^*$, the vector $\frac{1}{\theta}h$ can be generated by three vector-valued jointly Gaussian random variables. This implies that for three random variables the region of entropic vectors can be obtained by considering the cone generated by the space of Gaussian entropic vectors. It also suggests that studying Gaussian random variables for $n \geq 4$ may be a fruitful approach to studying the space of entropic vectors for arbitrary n .

I. INTRODUCTION

Obtaining the capacity region of information networks has long been an important open problem. It turns out that there is a fundamental connection between the entropy region of a number of random variables and the capacity region of networks [1] [2]. However determining the entropy region has proved to be an extremely difficult problem and there have been different approaches towards characterizing it. While most of the effort has been towards obtaining outer bounds for the entropy region by determining valid information inequalities [3], [4], [5], [6], [7], [8], [9], [10] some have focused on innerbounds [11], [12], [13] which may prove to be more useful since they yield achievable regions.

Let X_1, \dots, X_n be n jointly distributed discrete random variables with arbitrary alphabet size N . The vector of all the $2^n - 1$ joint entropies of these random variables is referred to as their “entropy vector” and conversely any $2^n - 1$ dimensional vector whose elements can be regarded as the joint entropies

This work was supported in part by the National Science Foundation under grant CCF-0729203, by the David and Lucille Packard Foundation and by Caltech’s Lee Center for Advanced Networking.

of some n random variables, for some alphabet size N , is called “entropic”. The *entropy region* is defined as the region of all possible entropic vectors and is denoted by Γ_n^* [3]. Let $\mathcal{N} = \{1, \dots, n\}$ and $\alpha, \beta \subseteq \mathcal{N}$. If we define $X_\alpha = \{X_i : i \in \alpha\}$ then it is well known that the joint entropies $H(X_\alpha)$ (or H_α , for simplicity) satisfy the following inequalities:

- 1) $H_\emptyset = 0$
- 2) For $\alpha \subseteq \beta$: $H_\alpha \leq H_\beta$
- 3) For any α, β : $H_{\alpha \cup \beta} + H_{\alpha \cap \beta} \leq H_\alpha + H_\beta$.

These are called the basic inequalities of Shannon information measures and the last one is referred to as the “submodularity property”. They all follow from the nonnegativity of the conditional mutual information [5], [14], [15]. Any inequality obtained from positive linear combinations of conditional mutual information is called a “Shannon-type” inequality. The space of all $2^n - 1$ dimensional vectors which only satisfy the Shannon inequalities is denoted by Γ_n . It has been shown that $\Gamma_2^* = \Gamma_2$ and $\bar{\Gamma}_3^* = \Gamma_3$ where $\bar{\Gamma}_3^*$ denotes the closure of Γ_3^* [5]. However, for $n \geq 4$, in 1998 the first non-Shannon type information inequality was discovered [5] which demonstrated that Γ_4^* is strictly smaller than Γ_4 . Since then many other non-Shannon type inequalities have been discovered [16], [6], [9], [10]. Nonetheless, the complete characterization of Γ_n^* for $n \geq 4$ remains open.

The effort to characterize the entropy region has focused on discrete random variables, ostensibly because the study of discrete random variables is simpler. However, continuous random variables are as important, where now for any collection of random variables X_α , with joint probability density function $f_{X_\alpha}(x_\alpha)$, the differential entropy is defined as

$$h_\alpha = - \int f_{X_\alpha}(x_\alpha) \log f_{X_\alpha}(x_\alpha) dx_\alpha. \quad (1)$$

Let $\sum_\alpha a_\alpha H_\alpha \geq 0$ be a valid discrete information inequality. This inequality is called *balanced* if for all $i \in \mathcal{N}$ we have $\sum_{\alpha: i \in \alpha} a_\alpha = 0$. Using this notion Chan [17] has shown a correspondence between discrete and continuous information inequalities, which allows us to compute the entropy region for one from the other.

Theorem 1 (Discrete/continuous information inequalities):

- 1) A linear continuous information inequality $\sum_\alpha a_\alpha h_\alpha \geq 0$ is valid if and only if its discrete counterpart $\sum_\alpha a_\alpha H_\alpha \geq 0$ is balanced and valid.

- 2) A linear discrete information inequality $\sum_{\alpha} a_{\alpha} H_{\alpha} \geq 0$ is valid if and only if it can be written as $\sum_{\alpha} \beta_{\alpha} h_{\alpha} + \sum_{i=1}^n r_i (h_{i,i^c} - h_{i^c})$ for some $r_i \geq 0$, where $\sum_{\alpha} \beta_{\alpha} h_{\alpha} \geq 0$ is a valid continuous information inequality (i^c denotes the complement of i in \mathcal{N}).

The above Theorem suggests that one can also study continuous random variables to determine Γ_n^* . Among all continuous random variables, the most natural ones to study first (for many of the reasons further described below) are Gaussians. This will be the main focus of this paper.

Let $X_1, \dots, X_n \in \mathcal{R}^T$ be n jointly distributed zero-mean¹ vector valued Gaussian random variables with covariance matrix $R \in \mathcal{R}^{nT \times nT}$. Clearly, R is symmetric, positive semidefinite, and consists of block matrices of size $T \times T$ (corresponding to each random variable). We will allow T to be arbitrary and will therefore consider the *normalized* joint entropy of any subset $\alpha \subseteq \mathcal{N}$ of these random variables

$$h_{\alpha} = \frac{1}{T} \cdot \frac{1}{2} \log \left((2\pi e)^{T|\alpha|} \det R_{\alpha} \right), \quad (2)$$

where $|\alpha|$ denotes the cardinality of the set α and R_{α} is the $|\alpha|T \times |\alpha|T$ matrix obtained by keeping those block rows and block columns of R that are indexed by α . Note that our normalization is by the dimensionality of the X_i , i.e., by T , and that we have used h to denote normalized entropy.

Normalization has the following important consequence.

Theorem 2 (Convexity of the region for h): The closure of the region of normalized Gaussian entropy vectors is convex. *Proof:* Let h^x and h^y be two normalized Gaussian entropy vectors. This means that the first corresponds to some collection of Gaussian random variables $X_i, \dots, X_n \in \mathcal{R}^{T_x}$ with the covariance matrix R^x , for some T_x , and the second to some other collection $Y_i, \dots, Y_n \in \mathcal{R}^{T_y}$ with the covariance matrix R^y , for some T_y . Now define the new set of random variables $Z_i = \left[(X_i^1)^t \dots (X_i^{N_x})^t (Y_i^1)^t \dots (Y_i^{N_y})^t \right]^t$ where $(\cdot)^t$ denotes the transpose, by stacking N_x and N_y independent copies of each, respectively, into a $N_x T_x + N_y T_y$ dimensional vector. Clearly the Z_i are jointly-Gaussian. Due to the independencies of the X_i^k and Y_i^l , $k = 1, \dots, N_x$, $l = 1, \dots, N_y$, the non-normalized entropy of the collection of random variables Z_{α} is

$$h_{\alpha}^z = N_x T_x h_{\alpha}^x + N_y T_y h_{\alpha}^y.$$

To obtain the normalized entropy we should divide by $N_x T_x + N_y T_y$

$$h_{\alpha}^z = \frac{N_x T_x}{N_x T_x + N_y T_y} h_{\alpha}^x + \frac{N_y T_y}{N_x T_x + N_y T_y} h_{\alpha}^y,$$

which, since N_x and N_y are arbitrary, implies that every vector that is a convex combination of h^x and h^y is entropic and generated by a Gaussian. ■

Note that h_{α} can also be written as follows:

$$h_{\alpha} = \frac{1}{2T} \log \det R_{\alpha} + \frac{1}{2|\alpha|} \log 2\pi e \quad (3)$$

¹Since differential entropy is invariant to shifts there is no point in assuming nonzero means for the X_i .

Therefore if we define

$$g_{\alpha} = \frac{1}{T} \log \det R_{\alpha}, \quad (4)$$

it is obvious that g_{α} can be obtained from h_{α} and vice versa. All that is involved is a scaling of the covariance matrix R . Therefore the space for g and h are the same. For simplicity, we will therefore use g_{α} instead of h_{α} throughout the paper and use the term entropy for both g and h interchangeably.

The remainder of this paper is organized as follows. In the next section we review some motivating results on the entropies of Gaussian random variables. Section III states the main results of the paper and Section IV outlines the proof.

II. SOME KNOWN RESULTS

One of the best known inner bounds for Γ_4^* is the so-called “Ingleton-bound” [12].

Theorem 3 (Ingleton inequality): [18] Let v_1, \dots, v_n be n vector subspaces and let $\mathcal{N} = \{1, \dots, n\}$. Further let $\alpha \subseteq \mathcal{N}$ and r_{α} be the rank function defined as the dimension of the subspace $\bigoplus_{i \in \alpha} v_i$. Then for any subsets $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \subseteq \mathcal{N}$, we have

$$r_{\alpha_1} + r_{\alpha_2} + r_{\alpha_1 \cup \alpha_2 \cup \alpha_3} + r_{\alpha_1 \cup \alpha_2 \cup \alpha_4} + r_{\alpha_3 \cup \alpha_4} - r_{\alpha_1 \cup \alpha_2} - r_{\alpha_1 \cup \alpha_3} - r_{\alpha_1 \cup \alpha_4} - r_{\alpha_2 \cup \alpha_3} - r_{\alpha_2 \cup \alpha_4} \leq 0 \quad (5)$$

The Ingleton inequality was first obtained for the rank of vector spaces. However it turns out that certain types of entropy functions, in particular all linear representable (corresponding to linear codes over finite fields) and pseudo-abelian group characterizable entropy functions also satisfy this inequality and hence fall into this inner bound [19], [20]. On the other hand if we consider 4 jointly Gaussian random variables, we interestingly find that they can violate the Ingleton bound. Consider the following covariance matrix:

$$\begin{bmatrix} 1 & \varepsilon & a & a \\ \varepsilon & 1 & a & a \\ a & a & 1 & 0 \\ a & a & 0 & 1 \end{bmatrix} \quad (6)$$

To violate the Ingleton inequality we need to have:

$$g_1 + g_2 + g_{123} + g_{124} + g_{34} - g_{12} - g_{13} - g_{14} - g_{23} - g_{24} \geq 0 \quad (7)$$

Substituting for values of g and simplifying we obtain:

$$\frac{1 - \varepsilon}{1 + \varepsilon} \geq \left(\frac{1 - 2a^2 + a^4}{1 - 2a^2 + \varepsilon} \right)^2 \quad (8)$$

Solving this inequality gives a region for permissible ε and a . In particular the point $\varepsilon = 0.25$ $a = 0.5$ lies in this region.

Interestingly enough, this example has also been discovered in a different context (the study of determinantal inequalities) in [21]. Moreover it has been shown that the Zhang-Yeung non-Shannon inequality [12] is tight for Gaussian random variables [22].

From (4) it can be easily seen that any valid information inequality for entropies can be immediately converted into an inequality for the (block) principal minors of a symmetric,

positive semi-definite matrix. Let g be the entropy vector corresponding to some vector-valued collection of random variables with an $nT \times nT$ covariance matrix R . Further, let m denote the vector of block principal minors of R . Then it is clear that $m = e^{gT}$, where the exponential acts component-wise on the entries of g . Then the submodularity of entropy translates to the following inequality for the principal minors:

$$m_{\alpha \cup \beta} \cdot m_{\alpha \cap \beta} \leq m_{\alpha} \cdot m_{\beta} \quad (9)$$

which is also known as the “Kotljanskii” inequality [21].

In passing we should mention that there has been interesting recent progress in determining the relationships between the principal minors of symmetric matrices [23], [24].

III. MAIN RESULTS

The above results (violation of the Ingleton bound and tightness of the non-Shannon inequality) lead one to speculate whether the entropy region for arbitrary continuous random variables can be obtained from the entropy region of (vector-valued) Gaussian ones. Unfortunately, the next theorem shows that this is not true for even $n = 3$.²

Theorem 4 (Entropy Region for $n = 3$ Gaussian RVs):

Consider the 7-dimensional vector $g = [g_1 \ g_2 \ g_3 \ g_{12} \ g_{23} \ g_{31} \ g_{123}]^t$ and define $x_k = e^{g_{ij} - g_i - g_j}$ and $y = \frac{\prod_k x_k}{\max_k x_k} + 2 \max_k x_k - \sum_k x_k$. Then g is a Gaussian entropic vector if, and only if,

1) If $y \leq 0$:

$$g_{ij} \leq g_i + g_j \quad , \quad g_{123} \leq \min_j (g_{ij} + g_{jk} - g_j). \quad (10)$$

2) And, if $y > 0$:

$$z = -2 + \sum_k x_k + 2 \sqrt{\prod_k (1 - x_k)} \geq 0 \quad (11)$$

$$g_{ij} \leq g_i + g_j \quad , \quad g_{123} \leq \sum_k g_k + \log z \quad (12)$$

The entropy region for three random variables is simply given by the inequalities in (10) above. Thus, when $y \leq 0$, the Gaussian entropy region coincides with the continuous entropy region; however, when $y > 0$ (and this can happen for some valid entropy vectors), we have the additional inequality (11) on the pairwise entropies, as well as the tighter upper bound (12) on g_{123} .

This implies that the Gaussian entropy region for $n = 3$ vector-valued random variables is *strictly* smaller than the actual entropy region.

Nonetheless, not all hope is lost and the next theorem shows that one can indeed construct the entropy region for $n = 3$ random variables from the entropy region generated by vector-valued Gaussians.

Theorem 5 (General and Gaussian Entropy Regions): Let $g \in \mathcal{R}^7$ be a continuous entropy vector. Then there exists

²The statement is, however, trivially true for $n = 2$ random variables, where the continuous entropy region is determined by the single inequality $g_{12} \leq g_1 + g_2$.

a $\theta^* > 0$, such that for all $\theta \geq \theta^*$, the vector $\frac{1}{\theta}g$ can be generated by three vector-valued jointly Gaussian random variables.

In other words, the entropy region for $n = 3$ continuous random variables is the (convex) cone generated by the entropy region of 3 Gaussian random variables. This result gives us hope that the study of Gaussians may be fruitful for $n \geq 4$.

IV. PROOF OF MAIN RESULTS

In what follows we will outline the proofs of Theorems 4 and 5 giving as much detail as space permits. The basic idea is to determine the structure of the Gaussian random variables that generate the *boundary* of the entropy region for Gaussians, and then to determine what the boundary entropies are.

Lemma 1 (Boundary of the Gaussian Entropy Region): The boundary of the Gaussian entropy region is generated by the concatenation of a set of vector valued Gaussian random variables with covariance

$$\begin{bmatrix} \alpha_{11}I_{\hat{T}} & \alpha_{12}\Phi_{12} & \alpha_{13}\Phi_{13} \\ \alpha_{12}\Phi_{12}^t & \alpha_{22}I_{\hat{T}} & \alpha_{23}\Phi_{23} \\ \alpha_{13}\Phi_{13}^t & \alpha_{23}\Phi_{23}^t & \alpha_{33}I_{\hat{T}} \end{bmatrix}, \quad (13)$$

where the Φ_{ij} are *orthogonal* matrices, and another set of independent vector-valued Gaussian random vectors with covariance

$$\begin{bmatrix} \alpha_{11}I_{T-\hat{T}} & 0 & 0 \\ 0 & \alpha_{22}I_{T-\hat{T}} & 0 \\ 0 & 0 & \alpha_{33}I_{T-\hat{T}} \end{bmatrix}. \quad (14)$$

Proof: To find the boundary region for 3 jointly Gaussian random variables, we should solve the following maximization problem:

$$\max_R \sum_{s \subseteq \{1,2,3\}} \gamma_s \log \det R_s, \quad (15)$$

where R is the $3T \times 3T$ block covariance matrix. This optimization comes about when we fix any 6 of the entropies and try to maximize the last one. KKT conditions then necessitate that for $i, j, k \in \{1, 2, 3\}$:

$$\left(\begin{bmatrix} \gamma_i R_{ii}^{-1} & 0 \\ 0 & \gamma_j R_{jj}^{-1} \end{bmatrix} + \gamma_{ij} \begin{bmatrix} R_{ii} R_{ij} \\ R_{ji} R_{jj} \end{bmatrix}^{-1} \right) \begin{bmatrix} R_{ik} \\ R_{jk} \end{bmatrix} = 0 \quad (16)$$

Moreover note that there is no loss of generality in assuming R_{ii} to be multiple of identity. In fact if $\det R_{ii} = \alpha_{ii}^T$ where $\alpha_{ii} \geq 0$, one can always pre and post multiply the covariance matrix R by the block diagonal matrix \hat{R} with elements $\hat{R}_{ii} = \sqrt{\alpha_{ii}} R_{ii}^{-1/2}$ to obtain $\alpha_{ii}I$ as the diagonal elements without changing any of the principal minors. Henceforth we will assume $R_{ii} = \alpha_{ii}I$.

Therefore simplifying condition (16), we obtain:

$$\begin{bmatrix} (\gamma_i + \gamma_{ij})I & \frac{\gamma_j}{\alpha_{jj}} R_{ij} \\ \frac{\gamma_i}{\alpha_{ii}} R_{ji} & (\gamma_j + \gamma_{ij})I \end{bmatrix} \begin{bmatrix} R_{ik} \\ R_{jk} \end{bmatrix} = 0 \quad (17)$$

Now if the $2T \times T$ matrix $\begin{bmatrix} R_{ik} & R_{jk} \end{bmatrix}^t$ were full rank, the rank of the left $2T \times 2T$ matrix would be T and therefore its Schur complement should be zero, i.e.:

$$(\gamma_j + \gamma_{ij})I - \frac{(\frac{\gamma_i}{\alpha_{ii}})(\frac{\gamma_j}{\alpha_{jj}})}{\gamma_i + \gamma_{ij}} R_{ji} R_{ij} = 0 \quad (18)$$

in other words:

$$R_{ji} R_{ij} = R_{ij} R_{ji} \triangleq \beta I \quad (19)$$

This implies that off-diagonal entries of R be multiples of an orthogonal matrix. However, in the general case $\begin{bmatrix} R_{ik} & R_{jk} \end{bmatrix}^t$ need not be full rank. Therefore there is a $T \times T$ unitary matrix θ_{ij} such that:

$$\begin{bmatrix} R_{ik} \\ R_{jk} \end{bmatrix} \theta_{ij} = \begin{bmatrix} \bar{R}_{ik} & 0 \\ \bar{R}_{jk} & 0 \end{bmatrix} \quad (20)$$

Assume the column rank of $\begin{bmatrix} \bar{R}_{ik} & \bar{R}_{jk} \end{bmatrix}^t$ to be T_{ij} . This suggests doing a similarity transformation on R with the following unitary matrix:

$$\Theta = \begin{bmatrix} \theta_{23} & 0 & 0 \\ 0 & \theta_{31} & 0 \\ 0 & 0 & \theta_{12} \end{bmatrix} \quad (21)$$

From which we obtain:

$$\Theta^* R \Theta = \begin{bmatrix} \alpha_{11} I & \theta_{23}^* R_{12} \theta_{31} & \theta_{23}^* R_{13} \theta_{12} \\ \theta_{31}^* R_{21} \theta_{23} & \alpha_{22} I & \theta_{31}^* R_{23} \theta_{12}^* \\ \theta_{12}^* R_{31} \theta_{23} & \theta_{12}^* R_{32} \theta_{31}^* & \alpha_{33} I \end{bmatrix} \quad (22)$$

Considering $R_{21} \theta_{23}$ and $\theta_{31}^* R_{21}$ simultaneously and using (20) we can simply obtain the following structure for $\theta_{31}^* R_{21} \theta_{23}$:

$$\theta_{31}^* R_{21} \theta_{23} = \begin{bmatrix} \hat{R}_{21} & 0 \\ 0 & 0 \end{bmatrix} \quad (23)$$

where the dimension of \hat{R}_{21} is $T_{31} \times T_{23}$. A similar argument for other elements yields the following structure for $\Theta^* R \Theta$:

$$\begin{bmatrix} \alpha_{11} I_{T_{23}} & 0 & \hat{R}_{12} & 0 & \hat{R}_{13} & 0 \\ 0 & \alpha_{11} I_{T-T_{23}} & 0 & 0 & 0 & 0 \\ \hat{R}_{21} & 0 & \alpha_{22} I_{T_{31}} & 0 & \hat{R}_{23} & 0 \\ 0 & 0 & 0 & \alpha_{22} I_{T-T_{31}} & 0 & 0 \\ \hat{R}_{31} & 0 & \hat{R}_{32} & 0 & \alpha_{33} I_{T_{12}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{33} I_{T-T_{12}} \end{bmatrix} \quad (24)$$

using (24) and plugging back into (17) we obtain:

$$\begin{bmatrix} (\gamma_i + \gamma_{ij}) I_{T_{jk}} & \frac{\gamma_j}{\alpha_{jj}} \hat{R}_{ij} \\ \frac{\gamma_i}{\alpha_{ii}} \hat{R}_{ji} & (\gamma_j + \gamma_{ij}) I_{T_{ki}} \end{bmatrix} \begin{bmatrix} \hat{R}_{ik} \\ \hat{R}_{jk} \end{bmatrix} = 0 \quad (25)$$

Note that the dimension of $\begin{bmatrix} \hat{R}_{ik} & \hat{R}_{jk} \end{bmatrix}^t$ is $(T_{jk} + T_{ki}) \times T_{ij}$ and that this equation is similar to (17) with \hat{R}_{ij} instead of R_{ij} . Therefore the same argument leads to the conclusion that \hat{R}_{ij} is a multiple of an orthogonal matrix say Φ_{ij} ; in other words $\hat{R}_{ij} = \alpha_{ij} \Phi_{ij}$. Moreover if we let the rank of the left matrix in (25) be r we will have:

$$r \leq T_{jk} + T_{ki} - T_{ij} \quad (26)$$

On the hand it is also obvious that:

$$r \geq T_{jk}, T_{ki} \quad (27)$$

From (26) and (27) it follows that:

$$T_{ij} \leq \min(T_{jk}, T_{ki}) \quad (28)$$

Since a similar argument can be used for T_{jk} and T_{ki} we conclude that:

$$T_{12} = T_{23} = T_{31} \triangleq \hat{T} \quad (29)$$

From which it follows that after a series of permutations $\Theta^* R \Theta$ can be written as follows:

$$\begin{bmatrix} \alpha_{11} I_{\hat{T}} & \alpha_{12} \Phi_{12} & \alpha_{13} \Phi_{13} & 0 & 0 & 0 \\ \alpha_{12} \Phi_{12}^t & \alpha_{22} I_{\hat{T}} & \alpha_{23} \Phi_{23} & 0 & 0 & 0 \\ \alpha_{13} \Phi_{13}^t & \alpha_{23} \Phi_{23}^t & \alpha_{33} I_{\hat{T}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{11} I_{T-\hat{T}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{22} I_{T-\hat{T}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{33} I_{T-\hat{T}} \end{bmatrix} \quad (30)$$

■ **Lemma 2 (Block Orthogonal, Block Diagonal Covariance):** Consider the covariance matrix

$$R = \begin{bmatrix} \alpha_{11} I_T & \alpha_{12} \Phi_{12} & \alpha_{13} \Phi_{13} \\ \alpha_{12} \Phi_{12}^t & \alpha_{22} I_T & \alpha_{23} \Phi_{23} \\ \alpha_{13} \Phi_{13}^t & \alpha_{23} \Phi_{23}^t & \alpha_{33} I_T \end{bmatrix}, \quad (31)$$

where the Φ_{ij} are orthogonal, $\alpha_{ii} > 0$, and the block principal minors $m_{ij} = p_{ij}^T = (\alpha_{ii} \alpha_{jj} - \alpha_{ij}^2)^T$ are such that $p_{ij} \geq 0$. Then

$$\det R \leq \left(\alpha_{11} \alpha_{22} \alpha_{33} - \alpha_{11} \alpha_{23}^2 - \alpha_{22} \alpha_{13}^2 - \alpha_{33} \alpha_{12}^2 + 2|\alpha_{12} \alpha_{13} \alpha_{23}| \right)^T \quad (32)$$

with equality iff $\Phi + \Phi^t = 2I$ where $\Phi = \Phi_{13}^t \Phi_{12} \Phi_{23}$.

Proof: It is not hard to show that,

$$\det R = \det \left((\alpha_{11} \alpha_{22} \alpha_{33} - \alpha_{11} \alpha_{23}^2 - \alpha_{22} \alpha_{13}^2 - \alpha_{33} \alpha_{12}^2) I_T + \alpha_{12} \alpha_{13} \alpha_{23} (\Phi + \Phi^t) \right) \quad (33)$$

The result immediately follows from $-2I \leq \Phi + \Phi^t \leq 2I$. ■

Lemma 3: Consider the function

$$f(\theta) = \max \left(0, -2 + \sum_{k=1}^3 x_k^{\frac{1}{\theta}} + 2 \sqrt{\prod_{k=1}^3 (1 - x_k^{\frac{1}{\theta}})} \right), \quad (34)$$

where $0 < x_k \leq 1$, for $k = 1, 2, 3$. f has a single maximum given by:

$$\max_{\theta} f(\theta) = \frac{\prod_k x_k}{\max_k(x_k)} \quad (35)$$

Moreover if we let $y = \frac{\prod_k x_k}{\max_k x_k} + 2 \max_k x_k - \sum_k x_k$,

1) If $y \leq 0$ then

$$\max_{0 \leq \theta \leq 1} f(\theta) = \frac{\prod_k x_k}{\max_k(x_k)} \quad (36)$$

2) if $y > 0$,

$$\max_{0 \leq \theta \leq 1} f(\theta) = f(1) \quad (37)$$

Proof: Omitted for brevity. ■

Now we can proceed to the proof of Theorem 4.

Proof of Theorem 4: To find the boundary entropies of the region we use Lemma (1) to time-share a set of independent random variables with covariance matrix of block size $T - \hat{T}$ and another set of random variables with orthogonal covariance matrix of block size \hat{T} (30). Calculating the determinant of this matrix and using Lemma (2) we obtain:

$$\max_{\Phi} \det R = (\alpha_{11}\alpha_{22}\alpha_{33})^{T-\hat{T}} \left(\alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}^2 - \alpha_{22}\alpha_{13}^2 - \alpha_{33}\alpha_{12}^2 + 2|\alpha_{12}\alpha_{13}\alpha_{23}| \right)^{\hat{T}} \quad (38)$$

Let m be the vector of block principal minors of the above matrix and let $m = p^T$ where the exponential acts componentwise. Then it is easy to see that $\alpha_{ii} = p_i \geq 0$ and $\alpha_{ij} = \pm\sqrt{p_i p_j - p_{ij}}$. This imposes the constraint:

$$p_{ij} \leq p_i p_j \quad (39)$$

Assuming $\theta = \frac{\hat{T}}{T}$ and substituting these in (38) result in:

$$\begin{aligned} \max_{\Phi} p_{123} &= p_1 p_2 p_3 \left\{ -2 + \left(\frac{p_{12}}{p_1 p_2} \right)^{\frac{1}{\theta}} + \left(\frac{p_{13}}{p_1 p_3} \right)^{\frac{1}{\theta}} + \left(\frac{p_{23}}{p_2 p_3} \right)^{\frac{1}{\theta}} \right. \\ &\quad \left. + 2 \sqrt{\left(1 - \left(\frac{p_{12}}{p_1 p_2} \right)^{\frac{1}{\theta}} \right) \left(1 - \left(\frac{p_{13}}{p_1 p_3} \right)^{\frac{1}{\theta}} \right) \left(1 - \left(\frac{p_{23}}{p_2 p_3} \right)^{\frac{1}{\theta}} \right)} \right\}^{\theta} \end{aligned} \quad (40)$$

Of course this corresponds to the determinant of a covariance matrix of some Gaussian random variables only if the term inside the braces in (40) is positive. Therefore assuming $x_1 = \frac{p_{12}}{p_1 p_2}$, $x_2 = \frac{p_{23}}{p_2 p_3}$, $x_3 = \frac{p_{13}}{p_3 p_1}$ and using (34):

$$\max_{\Phi, \theta} p_{123} = p_1 p_2 p_3 \max_{0 \leq \theta \leq 1} f(\theta) \quad (41)$$

what remains, is to maximize $f(\theta)$ with respect to θ over $0 \leq \theta \leq 1$. This is readily obtained using Lemma (3),

1) If $y \leq 0$,

$$\max_{\Phi, \theta} p_{123} = \min_j \frac{p_{ij} p_{jk}}{p_j} \quad (42)$$

2) If $y > 0$,

$$\max_{\Phi, \theta} p_{123} = \max \left(0, -2 + \sum_{k=1}^3 x_k + 2 \sqrt{\prod_{k=1}^3 (1 - x_k)} \right), \quad (43)$$

Defining $z = -2 + \sum_{k=1}^3 x_k + 2 \sqrt{\prod_{k=1}^3 (1 - x_k)}$ and insisting on z being positive yields (11). Replacing p with the corresponding entropies ($p = e^g$) in (42), (43) and also (39) completes the proof. ■

Proof of Theorem 5: Let g be an arbitrary entropy vector and let $\theta^* = \operatorname{argmax}_{\theta} f(\theta)$. Now if we define the normalized entropy vector $\tilde{g} = \frac{1}{\theta^*} g$, for any $\theta' \geq \theta^*$, the minors go to

the powers of $\frac{1}{\theta'}$ and from Lemma (3) it follows that the maximum of the corresponding function f will happen for some $0 \leq \theta \leq 1$ and therefore by Theorem (4) we conclude that $\tilde{g} = \frac{1}{\theta^*} g$ can be generated by Gaussians. ■

REFERENCES

- [1] B. Hassibi and S. Shadbakht, "Normalized entropy vectors, network information theory and convex optimization," in *Information theory workshop, Bergen, Norway*, 2007.
- [2] X. Yan, R.W. Yeung, and Z. Zhang, "The capacity region for multi-source multi-sink network coding," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2007, pp. 116–120.
- [3] R.W. Yeung, "A framework for linear information inequalities," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1924–1934, 1997.
- [4] F. Matus, "Infinitely many information inequalities," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2007, pp. 41–44.
- [5] Zhen Zhang and Raymond Yeung, "A non-shannon-type conditional inequality of information quantities," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1982–1986, 1997.
- [6] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-shannon information inequalities," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2006, pp. 233–236.
- [7] T.H. Chan and R.W. Yeung, "On a relation between information inequalities and group theory," *IEEE Trans. on Information Theory*, vol. 48, no. 7, pp. 1992–1995, 2002.
- [8] T.H. Chan, "A combinatorial approach to information inequalities," *Communications In Information and Systems*, vol. 1, no. 3, pp. 241–254, 2001.
- [9] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-shannon-type inequalities for entropies," *Communications In Information and Systems*, vol. 2, no. 2, pp. 147–166, 2002.
- [10] Z. Zhang, "On a new non-Shannon type information inequality," *Communications In Inf. and Systems*, vol. 3, no. 1, pp. 47–60, 2003.
- [11] B. Hassibi and S. Shadbakht, "On a construction of entropic vectors using lattice-generated distributions," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2007, pp. 501–505.
- [12] Z. Zhang and R. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. on Information Theory*, vol. 44, no. 4, pp. 1440–1452, 1998.
- [13] F. Matus and M. Studeny, "Conditional independences among four random variables I," *Combin., Prob. Comput.*, vol. 4, pp. 269–278, 1995.
- [14] S. Fujishige, "Polymatroidal dependence structure of a set of random variables," *Information and Control*, vol. 39, pp. 55–72, 1978.
- [15] T. S. Han, "A uniqueness of shannon's information distance and related nonnegativity problems," *J. Comb. Inform. Syst. Sci.*, vol. 6, no. 4, pp. 320–331, 1981.
- [16] F. Matus, "Two constructions on limits of entropy functions," *IEEE Trans. on Information Theory*, vol. 53, no. 1, pp. 320–330, 2007.
- [17] T. H. Chan, "Balanced information inequalities," *IEEE Trans. on Information Theory*, vol. 49, no. 12, pp. 3261–3267, 2003.
- [18] A. W. Ingleton, "Representation of matroids," in *Combinatorial mathematics and its applications*, D. Welsh, Ed. London: Academic Press, 1971, pp. 149–167.
- [19] T. H. Chan, "Capacity region for linear and Abelian network codes," in *Information theory and applications workshop, San Diego, CA*, 2007.
- [20] T. H. Chan, "Group characterizable entropy functions," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2007, pp. 506–510.
- [21] Charles R. Johnson and Wayne W. Barrett, "Determinantal inequalities for positive definite matrices," *Discrete Mathematics*, vol. 119, pp. 97–106, 1993.
- [22] Radim Lnenicka, "On the tightness of the Zhang-Yeung inequality for gaussian vectors," *Communications in information and systems*, vol. 3, no. 1, pp. 41–46, 2003.
- [23] O. Holtz and B. Sturmfels, "Hyperdeterminantal relations among symmetric principal minors," *Journal of Algebra*, vol. 316, pp. 634–648, 2007.
- [24] K. Griffin and Tsatsomeros M. J., "Principal minors, part II: The principal minor assignment problem," *Linear Algebra and its applications*, vol. 419, pp. 125–171, 2006.